Krieger, N., & Davey Smith, G. (2016). The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *International Journal of Epidemiology*, *45*(6), 1787-1808. https://doi.org/10.1093/ije/dyw114

Title The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology

Authors Nancy Krieger, PhD, Professor of Social Epidemiology, American Cancer Society Clinical Research Professor, Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health

George Davey Smith, MD, DSc, Professor of Clinical Epidemiology, MRC Integrative Epidemiology Unit at the University of Bristol

Corresponding author
Nancy Krieger, PhD
Professor of Social Epidemiology
American Cancer Society Clinical Research Professor
Department of Social and Behavioral Sciences, Kresge 717
Harvard T.H. Chan School of Public Health
677 Huntington Avenue
Boston, MA 02115 (USA)
office: 617-432-1571
fax: 617-432-3123
email: nkrieger@hsph.harvard.edu

Word count
abstract: 250
text: 6272

Tables 1
Textboxes 4
Figures 2

Submitted to *Int J Epidemiol* on April 7, 2015
Revised and resubmitted on February 20, 2016
Re-revised and resubmitted: March 24, 2016

Funding. NK was supported, in part, by her American Cancer Society Clinical Research Professor Award. GDS works within a Unit supported by the MRC (MC-UU-12013/6) and the University of Bristol.

Acknowledgments. Thanks to Debbie Lawlor, John Lynch, Caroline Relton, Kate Tilling, and Alexander Bird for insightful comments on an earlier draft of the paper.

Abstract

"Causal inference," in 21st c CE epidemiology, has notably come to stand for a specific approach, one focused primarily on counterfactual and potential outcome reasoning, and using particular representations, such as directed acyclic graphs (DAGs) and Bayesian causal nets. In this essay, we suggest that, in epidemiology, no one causal approach should drive the questions asked or delimit what counts as useful evidence. Robust causal inference instead comprises a complex narrative, created by scientists appraising, from diverse perspectives, different strands of evidence produced by myriad methods. DAGs can of course be useful, but should not alone wag the causal tale. To make our case, we first address key conceptual issues, after which we offer several concrete examples illustrating how the newly favored methods, despite their strengths, can also: (a) limit who and what may be deemed a "cause," thereby narrowing the scope of the field, and (b) lead to erroneous causal inference, especially if key biological and social assumptions about parameters are poorly conceived, thereby potentially causing harm. As an alternative, we propose that the field of epidemiology consider judicious use of the broad and flexible framework of "inference to the best explanation," an approach perhaps best developed by Peter Lipton, a philosopher of science who frequently employed epidemiologically-relevant examples. This stance requires not only that we be open to being pluralists about both causation and evidence but that we also rise to the challenge of forging explanations that, in Lipton's words, aspire to "scope, precision, mechanism, unification and simplicity."

Key messages (3-5 bullet points)

- Since the late 1990s, epidemiologic literature explicitly focused on causal inference, conceptually and methodologically, has burgeoned, with most of it employing counterfactual and potential outcome reasoning, to the point where the phrase "causal inference" is equated almost exclusively with "counterfactual causal inference," with formal representation encoded in directed acyclic graphs (DAGs).

- The 21st c CE epidemiologic emphasis on one approach to causal inference, however, stands in stark contrast to the equally recent explosion of literature on causal inference in philosophy and history of science, and also in diverse natural and social sciences, in which vibrant debates exist over types and processes of causal inference and explanation.

- In this essay, using the examples of pellagra, the "birthweight" paradox, and racism and health, we instead suggest that a more promising approach for epidemiology would be to consider judicious use of the broad and flexible framework of "inference to the best explanation" (IBE), which, in the words of the philosopher Peter Lipton, aims to "think through inferential problems in causal rather than logical terms," so as to reach what Lipton termed the "loveliest," and not just "likeliest," explanation, one characterized by "scope, precision, mechanism, unification, and simplicity."

- Methodologically, to strengthen causal inference and explanation, we underscore the need for causal triangulation, whereby epidemiologists should employ diverse study designs, each involving different and unrelated potential biases, and test our hypotheses in different populations and in different historical periods, to see if results are robust to the confounding structures encountered and the analytic methods used.

- DAGs and counterfactual approaches are but one set of conceptual tools that epidemiologists can employ, and should not occupy a privileged place in delimiting the kinds of questions we ask or causes we theorize.

Causal inference. These two words, knit together, have come to new prominence in contemporary epidemiology [1-10]. Whereas prior to 1990 not one article in the Web of Science was indexed with a title or "topic" pertaining to "causal AND inference AND epidemiology," as of the end of 2015, 558 such articles could be found, half of them published during or after 2010, with citations of these articles increasing exponentially (**Figure 1**)[11]. The stakes, after all, are high: riding on the findings of epidemiologic research are not only scientific credibility but also accountability and agency: who and what is shaping population distributions of health, disease, and well-being, within and across societies, and at what cost – and what benefit – to whom? [1,12-17]

Is it plausible to think, however, that epidemiologists did not concern themselves with inferring causation – and accountability – prior to 1990? Surely not. Insightful harbingers of today's debates were incisively developed in the final lengthy chapter – "In Search of Causes" – of Jerry Morris' classic 1957 text "*Uses of Epidemiology*" [13] – and received book-length treatment in Mervyn Susser's 1973 opus: "*Causal Thinking in the Health Sciences*" [14]. Disputes about elucidating causation likewise can be found in the epidemiologic literature of mid-20th c CE, e.g. in debates over tobacco [18-23], as well as in the mid-19th c CE, part and parcel of the emergence of population sciences [17,24-29].

In the epidemiology of the 21st c CE, however, "causal inference" is increasingly equated with one specific approach, which focuses primarily on counterfactual and potential outcome reasoning, and employs particular representations, such as directed acyclic graphs (DAGs) and Bayesian causal nets [1-4,6-10,30-35]. A key tenet is that the ability to discern (and quantify) "causal effects" hinges on positing counterfactuals that involve "manipulable" exposures which could in principle be randomized [1-4,6,7,30-35]. Indeed, in some expositions, "causal inference" has effectively become shorthand for "counterfactual causal inference" [1-4,6-10,30-35], as if no other approach to causal inference exists. Many (but not necessarily all) proponents of this approach further accept the premise that if an exposure cannot be "manipulated" (and, in effect, be randomized in principle, if not in actuality), it cannot produce "causal effects" [34-36].

These are strong claims. Not surprisingly, they are also contested: within and outside the field of

epidemiology [9,10,14,19,37-47]. Escalating debates about "causes," "causation," "evidence," and "explanations" are taking place in a wide variety of empirical population, policy, biological, and other natural sciences, and also disciplines that analyze science, e.g., philosophy, the history of science, and science and technology studies more broadly [37-50]. Within just the past 6 years, several large interdisciplinary tomes, each close to or exceeding 800 pages, have appeared, sporting such titles as: "*The Oxford Handbook of Causation*" (790 pages; 2009) [39], "*Causality in the Sciences*" (952 pages; 2011) [40], and "*Arguing About Science*" (795 pages; 2012) [38].

For epidemiology, a population science that necessarily straddles, simultaneously, the stochasticity (randomness) involved in the causes of individual cases and the population-level structuring of risk that produces predictable group-level differences [16,28,51,52], we argue that two issues are paramount. The first concerns who and what the current counterfactual framework of "causal inferences" is precluding from being deemed a "cause," thereby narrowing the scope of the field. The second focuses on how the newly favored methods, despite their undoubted strengths, can also potentially lead to spurious causal inference, especially if key biological and social assumptions about parameters are poorly conceived, thereby potentially causing harm.

In our view, no one causal approach should drive the questions asked or delimit what counts as useful evidence. Robust causal inference instead comprises a complex narrative, created by scientists appraising, from diverse perspectives, different strands of evidence produced by myriad methods [10,12-17,27,28,33,37-50]. DAGs can of course be useful [1-8,30-35], but should not alone wag the causal tale.

We argue instead that epidemiology, like any science, needs a flexible, multi-faceted, and historically-informed approach to causal inference. Only such an approach can grapple with the major complex public health issues of our times, among which are social inequalities in health within and between nations, the overlapping burdens of communicable and non-communicable diseases (including emerging infections), and the planetary emergency of environmental change, as driven by climate change [15-17,53-59]. To make our case, we first address key conceptual issues, after which we offer several concrete examples.

Clarifying causation: a case for pluralism about causes and types of evidence –and for explanation

Before jumping into the epidemiologic evidence, some clarifications are in order. First, we recognize that debates about what constitutes "causation" and demonstrating its existence have a long history – of at least a few millennia! [43] – and we obviously will not resolve these controversies in one essay. Second, our vantage is as pluralists: both about causality and about evidence [44,45], and we explain below what this entails. Third, our motivation to enter this debate is because we want to strengthen epidemiologic science and its capacity to contribute usefully to the multisectoral work urgently needed to improve population health and reduce, if not eliminate, health inequities [16,17].

In brief, within philosophical discourse the lack of a single theory or definition of "cause" is widely recognized, as is the notion that there is not just one method to identify causal processes and effects [37-46]. Two recent reviews, for example, have helpfully clarified [44,45] that not only are there "five families of `standard view' on causality" – i.e., "regularity, counterfactual, probabilistic, process/mechanist and agency/interventionist" [45, p. 769], but also that, for research conducted as guided by any of these "views," there also exists "evidential pluralism," referring to how "evidence of a variety of kinds – say, probabilistic, mechanistic, regularity – can bear on a causal hypothesis and strengthen it" [44, p. 27]. The implication is that "triangulation" of evidence "from a number of independent methods is one and perhaps the only way to be reasonably confident about the truth of the hypothesis" [44, p. 27].

Among the many reasons triangulation of evidence based on data from different contexts is important is recognition that the longer the causal "chain" or the larger the causal "network," the more likely that context-dependent effects are large enough to matter, implying the observed "effects" may be historically contingent [37-47]. Suggesting these are practical, not esoteric, concerns, UNAIDS in 2010 released a guide titled *An Introduction to Triangulation* [60], as part of their "monitoring and evaluation fundamentals" series. Intended to improve the monitoring and societal response to the HIV epidemic and other health outcomes, the booklet reviews the strengths and limitations of four widely-used types of

triangulation: "(1) data triangulation; (2) investigator triangulation; (3) theory triangulation; and (4) methodological or method triangulation" [60, p. 14] and further provides diverse empirical examples of why all four types of triangulation are necessary, since no one approach can guarantee robust causal inference. In our section on empirical examples, we provide concrete illustrations as to what such "triangulation" can entail for epidemiologic research.

Causal questions and answers, and hence inferences, may further depend on spatiotemporal scale and level [14,17,37-45]. Consider, after all, the classic question posed by the neurobiologist Steven Rose: what caused the frog to jump? [61, pp. 10-13; 62]. At the fast-and-tiny molecular level, an answer might be: the reaction of actin and myosin within a muscle cell. At the much slower and bigger level of organisms, an answer might be: the frog saw a snake and jumped in order to avoid being eaten. At the long-term and still larger level of species, still another answer might be: evolutionary processes leading to co-evolution of frogs and snakes as prey and predators in ecosystems affording niches for them both. Analytically distinct, all three answers are not only valid: they are concurrent, not sequential, inextricably embodied and joined in the instant that the frog jumps [62].

The same causal parsing applies to epidemiologic outcomes, as per the example of adiposity and cardiovascular mortality [63,64]. Thus, in a single instance, a death due to cardiovascular disease and cardiovascular mortality rates may be caused by individuals' adverse physiological and metabolic profiles (e.g., high blood pressure, high lipids) and by the sociopolitical and economic conditions that drive both the political economy of "Big Food" and population distributions of risk of weight gain and inadequate medical care [54,56,65,66]. Such a view expands options for different levels and types of preventive interventions. For persons already with high adiposity, population research at the molecular and physiological levels suggests that causal links between adiposity and risk of death due to ischemic heart disease can be alleviated, if not completely broken, by intervening, pharmacologically, physiologically, or through individuals' behavior changes, on such biological parameters as lipid profiles and blood pressure [67,68]. Additional research at the societal level points to the necessity of structural interventions to promote healthy ways of living, premised on conceptualization of food security and sustainability as a

human right, as opposed to treatment of food as primarily a for-profit commodity, so that all people can have access to affordable, nutritious, and pleasurable meals [54,65,66]. The point is both/and, not either/or.

Moreover, demonstrating that epidemiologists' concerns about narrow renderings of "causation" that omit societal causes is not new, **Textbox 1** presents an analogous "fable," published shortly after the end of World War I by the epidemiologist F.G. Crookshank (1873-1933), in an essay titled "First principles: and epidemiology," in which a single-minded police surgeon avers that if the cause of death by murder is a bullet, then the cause of death by war is many bullets (and sometimes also poisonous gas) [69]. To Crookshank, it was ludicrous to posit that germs alone were the single "true cause" ("*causa vera*") of epidemics, and the only legitimate target of both inference and intervention; instead, for both war and epidemics, there could be no avoiding of discussion of "racial, economic, or political conditions," not simply as "predisposing factors" but as causes in their own right [69, pp. 178-179]. Social and political challenges to vaccine distribution, e.g. for polio, measles, and human papilloma virus (HPV), serve to underscore this point [70-73].

Of course, as with counterfactuals, the danger lies in where one draws the line, to avoid infinite regress as to the number of factors that need be considered. Continuing the military metaphor, **Figure 2** shows an alarming example of the ultimate arrow salad – or spaghetti: a PowerPoint slide prepared in 2009 about US military strategy in Afghanistan [74,75]. Even Crookshank might have been daunted.

Causal judgments: inference to the best explanation

Fortunately, recent work on "inference to the best explanation" (IBE), especially as articulated by the philosopher Peter Lipton (1954-2007) [42,76,77], can provide epidemiologists – and other scientists – with an alternative cogent, historically-grounded, conceptual approach to thinking about, sorting through, and arriving at robust explanations [5,42,45,48,49,78-80]. Curiously, although epidemiological research has been integral to Lipton's arguments – as per his analysis of Ignaz Semmelweis's 1844-1848 research on childbed fever [42, pp. 74-90] (see **Textbox 2**), discussion of IBE in the epidemiologic literature is surprisingly limited [5,48,49,80] – and nowhere to be found in many leading epidemiologic publications

on causal inference [1-4,6,7,30,32,35].

What, then, is IBE? As explained by Lipton and other philosophers of science, IBE is a type of reasoning widely used by scientists (and most people in everyday life) [42,76-79]. It is also increasingly viewed by philosophers and historians of science as being, in the words of Douven, the "cornerstone of scientific methodology" and also "medical diagnosis" [78], with the latter notably and necessarily requiring cross-level inferences bridging from knowledge about unique individual patients to group-level regularities [81]. IBE's primary concern is *explanation*, an expansive task that requires critically reasoning about extant (and missing) evidence and competing hypotheses that could explain the evidence. Reliant on one type of inductive reasoning, variously termed "abduction" or "defeasible" reasoning (see **Table 1** for definitions) [42,78,79,82], IBE does not and cannot afford the same pristine certainty provided by deductive reasoning, whereby the conclusion logically must be true if the premises are true (e.g., Sam is a person, all people are mortal, therefore Sam is mortal). Though this might seem a drawback, contemporary scholarship increasingly demonstrates that IBE far better reflects the actual practice of science, advances in scientific explanation, and the successful implementation of what has been learned, in such diverse fields as the physical, biological, epidemiological, clinical, or social sciences, as compared to the idealized hypothetico-deductive approach [37-42,76-79,82], which over the past 30 years has been variously lauded [83-86], rejected [87,88], and accepted in modified form [89,90] in the epidemiological literature.

In brief, the essence of the IBE approach is to "think through inferential problems in causal rather than logical terms" [42, p. 208] and to employ a "two-stage mechanism involving the generation of candidate hypotheses and then selection from among them" [42, p. 208]. IBE is thus driven by theory, substantive knowledge, and evidence, as opposed to being driven solely by logic or by probabilities. Absolute Cartesian skepticism is rendered moot, since the emphasis is on best explanation, as opposed to the conjuring of any explanation, however improbable (or useless) [78]. Nor is IBE hobbled by a common problem that deductive reasoning cannot resolve: how to evaluate competing hypotheses when none are logically refuted by the extant evidence [76, pp. 452-453]; for examples, see **Textbox 2** regarding

Lipton's analysis of how Semmelweiss adjudicated between such competing hypotheses regarding cause(s) of childbed fever.

Guiding choice among explanations for IBE is a contrastive approach geared to identifying what Lipton has termed the "loveliest" as opposed to merely "likeliest" hypothesis, whereby criterion for "loveliest" include: "scope, precision, mechanism, unification and simplicity" [76, p. 423]; "prediction" does not garner special consideration because opposing hypotheses may still both predict a given phenomenon (e.g., disease rates higher in groups exposed vs. not exposed to X), but not be equally "lovely." Moreover, by emphasizing the need to test aptly chosen contrastive hypotheses, the IBE approach (per the examples provided in **Textbox 2** for childbed fever) provides guidance for explanatory causal reasoning that goes beyond listing whether the evidence is, minimally, coherent (as per the Hill criteria) [42,76,77].

IBE is additionally highly attuned to contextual knowledge, and hence to the claims involved when assertions are made about "all else being equal" – whether via experimental design or statistical "control" [46,47,78,79]. It thus underscores the inevitable reliance, for good or for bad, upon scientific judgment. From the standpoint of IBE, "causal inference" cannot be reduced to what the philosopher Stathis Psillos has termed "topic-neutral and context-insensitive" algorithms [79, p. 441], whether involving deductive logic or Bayesian statistics. Core to IBE is the understanding that there are no clear-cut rules or short cuts that minimize the need to amass substantive expertise and to generate and think critically about contrastive hypotheses – but nor is it the case that "anything goes."

Stated another way, IBE clarifies that data never speak by themselves – either to computer algorithms or to people – nor do beliefs about probabilities simply drop from the sky. Active scientific judgment is inevitably involved, with regard to who and what is included and excluded. Scientists accordingly are enjoined to think about the full range of evidence, not just data germane to one specific hypothesis, and also to test hypotheses with diverse sets of methods whose assumptions are uncorrelated, so as to strengthen causal inference [5,42,45], a point we discuss further in relation to the empirical examples we next analyze. Although epidemiologists have long been aware of the need to compare data

across the proverbial "time, place, and person" [5,12-14,91] (or, rather, social group [17]), the emphasis on comparison across methods and causal inference frameworks is more recent [5,42,45].

IBE further points to the necessity of eschewing the hubris of assuming that scientists can exhaustively delineate the profound complexity and quirkiness of the biophysical and social worlds in which we live, a world in which unanticipated discoveries of unimagined phenomena and causal connections are as much the rule as they are the exception [16,17,37-42,47,76,77]. One would need infinite knowledge, after all, to generate an exhaustive list of all conditions or factors that would ensure such assumptions as "other things being equal" or "other things being absent." Who would have thought, for example, prior to work conducted in the past decade, that olfactory receptors, in both humans and other species, occur in just about every organ, including our skin, and are not just restricted to the nasal passage? [92,93] While the issue is far from closed, an explanatory reframing of these receptors as specialized evolved chemical detectors, not solely for smell, notably opens up a previously untheorized biological possibility, one with potential epidemiological as well as clinical relevance. An analogous case, relevant to cancer and cardiovascular disease, has been the explanatory reframing of estrogen from being a molecule primarily or solely preoccupied with "sex" and reproductive tissues to being a steroid involved in cell growth and apoptosis throughout the body [94,95], with the expression of estrogen receptors being both tissue-wide and highly responsive to exogenous stimuli [96-98]. Different conceptualizations of key parameters and different explanations entail different scientific programs and different interventions, one of the many reasons that debates over causal inference are so charged.

Seeking explanations: epidemiological examples

We now redirect our focus to three concrete epidemiologic examples. Our purpose is to show why we cannot restrict the work of causal inference to solely a counterfactual approach, and why we may well do better to rise to the challenge of attempting to infer the best explanation.

What a DAG cannot discern: the case of pellagra

We start with a seemingly simple yet informative example: explaining why rates of pellagra were

high among children in the US south who were institutionalized in orphanages in the early 20th c CE, as compared to other children in the region who were not institutionalized [99-104]. During this period, major debates, within and across causal levels, raged over whether pellagra – a disease whose prevalence was known to be both high and seasonal among people whose diet was primarily based on corn – was caused by an infectious agent, a fungus, stress, heredity, or even capitalism itself [99-105].

Why the association between institutionalization and the disease? The two leading hypotheses involved the same causal elements, but the arrows pointed in entirely opposite directions. The "germ theory" hypothesis held that children who came to orphanages had a higher rate of infection, which they then more readily transmitted to other children within the crowded orphanages (but then: why did the staff not also get ill?). The contaminated food hypothesis held that the institutions caused the higher rates of pellagra because they served tainted food, i.e., contaminated corn mush (but then why did staff, who sometimes also ate the corn mush, not get ill?) [101]. The "stress" and "capitalism" hypotheses [99,100], while perhaps accurately identifying causes and aspects of the plight of institutionalized children, nevertheless did not explain why institutionalized impoverished children everywhere did not get pellagra. Both hypotheses could be represented by a DAG including the same elements, but with causal arrows in the reverse direction.

To resolve these conundrums, Joseph Goldberger devised an entirely new hypothesis: institutions caused the higher rates of pellagra because they served *deficient* food, whereby the orphanages fed children a poverty diet of corn mush supplemented by little else (whereas the staff ate not only the corn mush but also other more nutritious food, thereby preventing pellagra) [99-104] – and he conducted experiments with people (including himself, relatives, colleagues, and prisoners) and animals to test his hypothesis [99-104,106,107]. Later research revealed the missing factor was niacin, i.e., Vitamin B3 [104].

Of note, Goldberger's hypothesis used the same three key variables ("orphans," "institutions," "pellagra") employed in the two dominant rival hypotheses ("germ" and "contamination") – but utterly transformed understanding of the causal relationships at play by introducing to the equation what was

then a new way of thinking about etiology: disease arising from deficiency, not excess. His alternative hypothesizing thus would yield a DAG with the same anchoring elements but totally different causal pathways, reflecting a new understanding of mechanisms of disease causation. Goldberger's hypothesis, initially ill-received and unlikely, thus had to battle for recognition – and among the three, it was also, in Lipton's terminology, the "loveliest."

Why? Because, as Goldberger and his colleague Edgar Sydenstricker [12] emphasized at the time [108-110], it explained not only: (a) who did and did not contract pellagra at the orphanages, (b) the seasonal nature of the disease (as tied to when money for varied foods ran out, after the harvest season, among impoverished sharecroppers in the US south, leading to a diet of primarily corn mush leavened by some pork fat and perhaps a few greens) but also (c) why the disease was so common in the US South among impoverished (and/or institutionalized) persons, but was not so common among impoverished (and/or institutionalized) persons in the US North (because the former relied far more heavily than the latter on corn-mush diets) [99,106,108-110].

Granted, Goldberger's hypothesis was not popular among US southern politicians or public health officials [99-104]. Why? Because it placed blame on not only the orphanages but also the structural institutions that protected sharecropping and high rates of southern poverty [99,100,110]. To Goldberger and Sydenstricker, however, understanding the interplay of causes across and within levels was essential for effective action in public health [12,99,100,110] – a truncated account would not suffice.

One final useful point raised by the example of pellagra concerns why technical manipulability should not be confused with causal powers. Thus, whether or not people had the technology to isolate and manipulate levels of Vitamin B3, its absence and presence still produced causal effects. Nevertheless, using a mixture of observational and experimental epidemiological evidence, along with a hefty dose of theorizing informed by deep knowledge of infectious disease epidemiology, Goldberger was able to arrive at a pragmatic causal explanation that, using Susser's causal lexicography, got it "right enough" [14] to enable important effective preventive interventions to be implemented [99-104,106-110].

When is a methodological solution not the answer: using biology – and "triangulation" – to parse the puzzle of smoking, infant mortality, and the "birthweight paradox"

Next, we consider an example where it may be that the reasoning encoded in DAGs may have initially appeared to solve a paradox, only for further work to clarify that the proposed solution potentially may not be a satisfactory – or indeed "lovely" – deep explanation. The case is that of the well-known "birthweight paradox," which first garnered attention in the early 1960s, as part of the disputes (fueled by tobacco company funding [22,23]) over whether smoking harms health [111,112]. In brief, the apparent paradox was then (and remains now) that although the on-average birthweight is *lower* for liveborn infants exposed vs. not exposed to tobacco smoke as fetuses, nevertheless the infant mortality rate among low birthweight infants is *higher* among infants unexposed vs. exposed to tobacco smoke when in utero [111-114]. The counterintuitive implication is that maternal smoking is protective for infant mortality for liveborn low birthweight infants.

Over the past 40 years, many rounds of arguments have appeared in the pages of many journals, offering diverse appraisals as to whether the "paradox" is "real," as opposed to an artifact created by selection bias, choice of wrong referent or "at risk' groups (e.g. fetus vs. liveborn infant), etc. [111-118]. As interest in using DAGs in epidemiology began to rise in the early 21st c CE, this "paradox" not surprisingly presented itself as a ripe candidate for analysis. The first round of papers using DAGs to address this paradox generally concluded that "collider bias," i.e., introduction of confounding by an unmeasured factor due to conditioning on intermediate factor (in this case, smoking), is the cause of the apparent "paradox" [114,116,117]. The take-home message of these papers is that the paradox is resolved: the problem has been dealt with by appropriate methods. In other words, the explanation is to explain away the observed association as a consequence of bias induced by faulty methods.

But is this apparent end of the story? Suggesting there may be yet more wags to this particular tale, an elaborated and biologically plausible alternative explanation exists, one that may well do a better job at being "lovelier," by virtue of elucidating mechanisms and opening up possibilities for unifying understanding of other seemingly unrelated "paradoxes." It is that infants who are low birthweight for

reasons other than smoking may well have experienced harms during their fetal development unrelated to and much worse than those imposed by smoking, e.g., stochastic semi-disasters that knock down birthweight, as a result of random genetic or epigenetic anomalies affecting the sperm or egg prior to conception or arising during fertilization and embryogenesis [16,114,118].

Of note, the proposed alternative biological explanation cannot be discerned from a DAG. Indeed, as pointed out in a new reflection on using DAGs to parse this paradox, the DAGs for collider bias and for heterogeneity of low birthweight phenotypes have an identical structure [117]. A larger and "lovelier" point is that profoundly different causal pathways can result in two distinct groups nevertheless exhibiting the same state – and a DAG, by itself, cannot resolve which hypothesized pathways, if any, are correct.

An IBE approach further recognizes that no one study design can provide a definitive robust test of the hypotheses at issue. Instead, as noted above, what is required is evidential pluralism, i.e., triangulation of evidence from empirical studies whose methodological assumptions, limitations, biases, and errors (which inevitably affect all studies) are uncorrelated [5,42-45,48,49,60,76,77,119]. In **Textbox 3** we provide examples of what such systematic triangulation of evidence, derived using approaches with different biases, entails for the example of smoking and birthweight [119-125].

A similarly structured paradox, also involving children's health, generated even more heated discussion 70 years before the birthweight paradox – and likewise demonstrates the important value of the type of reasoning encoded in DAGs and also the work needed to determine if the underlying encoded assumptions are biologically and socially sound. In 1910, Karl Pearson and colleagues reported data apparently showing no detrimental effects of parental alcoholism on the health and development of their offspring [126,127], results which not surprisingly generated fierce controversy [128]. The economist A.C. Pigou, in an elegant riposte, pointed out how selection of the sample could generate such a null association even when an adverse influence existed in the overall population [129], thereby describing what would today be termed "collider bias" [130].

Pigou's description of how this seeming paradox could arise was specific to the particular conditions of Pearson's investigations [129]. Attesting to the value of DAGs for identifying the

"transportability" of the identified type of bias, i.e., the conditions under which it can affect other investigations [33], other similarly structured explanations for particular issues have been produced in the epidemiologic literature many times since, from Berkson's presentation of what became his eponymous bias [131] through Greenland and Neutra's discussion of a potentially misleading study design proposed to investigate the influence of endogenous estrogens on endometrial cancer [132]. As these examples suggest, formal formulation of such potential biases – which can be represented in DAGs – clearly provides an incisive way of extending thinking about bias from one situation to another, one that can aid the overall evaluation of evidence in any given particular situation.

It is another matter entirely, however, to elucidate, empirically, if the hypothesized biases do indeed exist and if they are sufficient to generate the observed associations. At issue is not simply whether a potential bias exists, but whether the plausible magnitude of its quantitative effect is sufficient to bias the study results [133,134].

Nor can a DAG provide insight into what omitted variables might be important or whether a variable is even conceptualized appropriately (as per the pellagra example); only use of relevant scientific theories (including epidemiologic theories of disease distribution) can aid conceptualizing the phenomena that co-produce the hypothesized causal relationships [13-17,20,21,37-43]. A corollary is that despite the clear value of DAGs for formalizing certain types of biases, this feature does not mean this approach has more inferential value compared to components of evidence that cannot be disciplined in this way, e.g., the structuring effects of macroeconomic and social forces. An appeal for "evidence-based" policies that relies, say, solely on randomized clinical trials or other interventions carried out on individuals will inevitably lead to debased policy making, as we have argued elsewhere [135,136]. Causes do not cease being causes if they are challenging to study.

Racism and health: the harm caused by spurious "causal inference" and "counterfactuals"

Our final example accordingly concerns a structural determinant, using the long-argued case of racism and health [137-141]. One alarming feature of late 20th and current 21st c CE epidemiologic

literature on "causal inference" is the re-appearance of previously rebutted causal claims that "race" is an individual "attribute" and that it cannot be a "cause" because is not "modifiable" [1,34,36,142-145]. Five such examples are provided in **Textbox 4**, culled from diverse public health, epidemiologic, biostatistics, and sociological publications [1,34,142-144]. They are congruent with new lines of contested work, supported by considerable NIH funding, that seek to "re-molecularize" race [146-153].

However, we clarified back in 2000 [50] and reiterated since [154], in accord with a considerable literature extending back to the 19th c CE [17,137-141,147,152-162], the problem – one with enormously harmful public health and policy implications -- is that this approach to causal inference and counterfactuals starts at the wrong level, and uses DAGs to bark at the wrong tree, and indeed miss the forest entirely.

What is the problem with viewing "race" as an "inherent feature of individuals" [1, p. 70] or as an "immutable characteristic" [144, p. 775]? The problems are two-fold: bad biology and bad social science, compounded by an ahistorical approach to both the literature and the evidence. First, with regard to bad biology, this belief fails to acknowledge reams of genetic evidence demonstrating *H. sapiens* cannot meaningfully be parsed (including by so-called "cluster" programs) into discrete genetically distinct "races" who can be singularly identified by a set of traits and for whom variation within groups is less than variation between groups [146-153]. By now, the notion of discrete, let alone "fixed," "races," especially in countries such as the US, with its history of being a colonial-settler and immigrant nation that also imported slaves and upheld legal slavery for centuries (1619-1865) , is especially absurd [17,28,54,146-153].

Second, with regard to bad social science, the view of "race" as, in effect, a "natural" kind (existing *a priori*, as a "real" grouping that exists independent of human classificatory schemes), completely disregards nearly two centuries' worth of scholarship on the histories of the social creation – and enforcement, by law, by force, and by terror – of the varied "racial" categories deployed in diverse societies, let alone their changing permutations over time [17,28,54,137-141,146-156,162-168]. It also ignores how these "racial" categories, like any social relationship, are co-constitutive: each is defined and

bounded in relation to the other, just as are master and slave, and masculine and feminine [28,137, 169,170]. Change the social relationship, and the categories and how people relate to them and what they mean for their lives and their health will consequently change as well. This type of dynamic co-causation, replete with feedback loops, however, is not what is conventionally (or easily) depicted in DAGs.

Even so, epidemiologic evidence provides supportive evidence for the hypothesis that modification of race relations causes changes in the population distributions of health. The relevant counterfactual pertains to racism, not "race." Examples include studies showing the beneficial impact of the abolition of Jim Crow in the mid-1960s on black/white inequities in infant mortality rates, above and beyond improvements linked to such Great Society programs as the "War on Poverty" and the introduction of Medicare, Medicaid, and desegregated health care facilities [172-176]. Causing these "modifications" was the power of social movements, which challenged structural racism, forced repeal of unjust laws, and created space and resources for health and social scientists and health and social work practitioners to provide input into newly possible programs [177-180]. Treat "race" as a given and focus only on discrete "factors," such as "income," as some proponents of the DAG approach propose [34,142,145], and a DAG will tell a biased tale that is woefully incomplete for guiding policy and promoting health equity. Although such realities do not sit easily with admonitions for epidemiologists to focus only on "causes" that can be "modified" by health or policy professionals [181-183], they are the facts we confront when dealing with health inequities.

The larger implication is that the "loveliest" explanation of racial/ethnic inequities in health is the one that engages most deeply with the ugly social facts of past and present realities of racial inequality and its myriad social, economic, and embodied manifestations [17,62,137-141]. Far more comprehensive explanations of the epidemiologic evidence can be achieved if, rather than treat "race" as an unmodifiable "inherent feature" and posit either an endless and illusory set of "racial" genetic differences in gene frequency for each and every ailment or a set of material conditions that are held to be "modifiable" without addressing inequitable race relations, we instead tackle the causal relationships between racism and health head on. To do so, we can be aided by the central insights of the ecosocial theory of disease

distribution and focus attention on how people literally embody, biologically, their societal and ecological context, thereby producing population patterns of health, disease, and well-being [17,62,184,185].

Conclusion

We deliberately have not offered one prescription for how epidemiologists can best infer causation. No such prescription exists. Nor, of course, do we suggest that some approaches (e.g., use of DAGs where appropriate) be ruled out of court and banished to the dog-house. Instead, as we hope the examples we have provided demonstrate, there is no short cut for hard thinking about the biological and social realities and processes that jointly create the phenomena we epidemiologists seek to explain, always with an eye towards producing knowledge that we and others can use to improve population health, reduce preventable suffering, and, we add, advance health equity.

To accomplish these goals, we advocate that the field of epidemiology consider judicious use of the broad and flexible framework of "inference to the best explanation." This stance requires not only that we be open to being pluralists about both causation and evidence but that we also rise to the challenge of forging explanations that aspire to "scope, precision, mechanism, unification and simplicity" [42, p. 423].

No single study, however beautifully designed, can unequivocally demonstrate causation. To improve our causal explanations, we would do best instead to opt for causal triangulation [5;42, pp. 53-54; 44, p. 27;48,49,60,186,187]. In practical terms, as illustrated by **Textbox 3,** this means systematically employing and assessing evidence in relation to diverse study designs, involving different methodological assumptions and biases [5,48,49,60,186,187], and also testing our hypotheses in different populations and in different historical periods [5,60,186-190], to see if results are robust to the confounding structures encountered and the analytic methods used. In essence, the biases for each method employed – since, of course, all methods have potential biases – would be through different processes, and unrelated, to the biases in the other methods. DAGs and counterfactual approaches are but one set of conceptual tools that epidemiologists can employ, and should not occupy a privileged place in delimiting the kinds of questions we ask or causes we theorize.

We would hazard the guess that many who advocate these styles of thought would probably agree with our position, but might not see the current emphasis on applying formal rules as leading to questions becoming restricted to those which fit neatly within these rules. Suggesting, however, that we are not raising straw arguments are narrow framings of what constitutes legitimate causal inference accompanying the burgeoning use of these methods and advocacy to do so [1,6-8,11,30-32,34,35,142-145]. Our fear is that these new "cutting-edge" methods will, by virtue of their rule-bound nature, limit the scope of epidemiology and its impact on the urgent real-world problems of global population health [9,10,17,33,53-56].

We close by noting that in 1957, Jerry Morris included in "*Uses of Epidemiology*" a section he titled "Changing People in a Changing Society," in which he raised a series of question that have fruitfully shaped the field's research program for now well over a half century [13, pp. 19-23]. Among his many questions were [13, p. 22]:

> "What are the implications to Public Health of more married women going out to work? And less of the older men? Of still increasing urban – and suburbanization? The rapid growth of new towns? Smokeless zones (still with sulphur)? The building of new power stations? Of less physical activity in work and more bodily sloth generally? … Of the more than 1000 extra motor vehicles a day? Of the rising consumption of sugar … Of the cheapening of fats? … Such questions (of contemporary history, it might be said) could readily be multiplied."

Noting that "[s]ome of the issues mentioned above cannot yet be framed in scientific terms; but parts at least of others could be tackled more energetically," Morris nonetheless optimistically averred: "Perhaps epidemiology with its special skills in identifying what matters more and what matters less, its concern for woods rather than trees, perhaps the epidemiological method can simplify such issues and usefully raise some bold questions about these too" [13, p. 23].

Any approach to causal inference that cannot help us answer the kinds of prevention-oriented questions that Morris posed, that cannot brook analysis of inequitable social relations as a cause of population health and health inequities [50-56,137-141], and that cannot conceive how to address the causal epidemiologic implications of the planetary crisis of global climate change [57-59,191,192], is inadequate – and if it restricts what questions can be asked, it is wrong. We can – and must – do better.

REFERENCES

1. Glass TA, Goodman SN, Hernán MA, Samet JM. Causal inference in public health. *Annu Rev Public Health* 2013; 34:61-75.
2. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiological research. *Epidemiol* 1999; 10:37-48.
3. Robins JM, Hernán MA, Brumbach B. Marginal structural models and causal inference in epidemiology. *Epidemiol* 2000; 11:550-560.
4. VanderWeele TJ, Vansteelandt S, Robins JM. Marginal structural models for sufficient cause interactions. *Am J Epidemiol* 2010; 1717:506-514.
5. Richmond RC, Al-Amin A, Davey Smith G, Relton C. Approaches for drawing causal inferences from epidemiological birth cohorts: a review. *Early Human Devel* 2014; 90:769-780.
6. VanderWeele T. *Explanation in causal inference: methods for mediation and interaction*. New York: Oxford University Press, 2015.
7. Hernán MA, Robins JM. *Causal Inference*. Boca Raton: Chapman & Hall/CRC, forthcoming (2016). Available at: http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/ ; accessed: March 23, 2016.
8. Porta M, Vineis P, Bolúmar F. The current deconstruction of paradoxes: one sign of the ongoing methodological "revolution." *Eur J Epidemiol* 2015; 1-9. [Epub ahead of print]
9. Broadbent A. Causation and prediction in epidemiology: a guide to the "methodological revolution." *Studies Hist Philosophy Biol Biomedical Sci* 2015; 54:72-80.
10. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol* 2016: dyv341.
11. Web of Science. Available at: http://apps.webofknowledge.com.ezp-prod1.hul.harvard.edu/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=1CMnQV8qjBQMk8nIloV&preferencesSaved= ; accessed: January 28, 2016.
12. Sydenstricker E. *Health & Environment*. New York: McGraw Hill, 1933.
13. Morris JN. *Uses of Epidemiology*. Edinburgh: E. & S. Livingstone, Ltd, 1957.
14. Susser M. *Causal Thinking in Health Sciences: Concepts and Strategies of Epidemiology*. New York: Oxford University Press, 1973.
15. Davey Smith G, Ebrahim S. Epidemiology – is it time to call it a day? *Int J Epidemiol* 2001; 30:1-11.
16. Davey Smith G. Epidemiology, epigenetics and the `Gloomy Prospect': embracing randomness in population health research and practice. *Int J Epidemiol* 2011; 40:537-562.
17. Krieger N. *Epidemiology & The People's Health: Theory & Context*. New York: Oxford University Press, 2011.
18. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. *JNCI* 1959; 22:173–203 (Reprinted *Int J Epidemiol* 2009; 38:1175-1191).
19. Stallones, RA. *The Association between Tobacco Smoking and Coronary Heart Disease. Draft report of June 28 [1963] to the Surgeon General's Advisory Committee on Smoking and Health*. *Int J Epidemiol* 2015; 44:735-743.

20. Blackburn H. Commentary: "The association between tobacco smoking and coronary heart disease." An unpublished memorandum by Reuel Stallones, consultant in cardiovascular diseases to the U.S. Surgeon General's Advisory Committee on Smoking and Health, 28 June and 19 July 1963. *Int J Epidemiol* 2015; 44:744-748.

21. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; 58:295-300.

22. Brandt AM. *The Cigarette Century: The Rise, Fall, and Deadly Persistence of the Product that Defined America*. New York: Basic Books, 2007.

23. Proctor R. *Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition*. Berkeley, CA: University of California Press, 2011.

24. Yeo EJ. Social surveys in the eighteenth and nineteenth centuries. In: Porter RM, Ross D (eds). *The Cambridge History of Science, Vol 7: The Modern Social Sciences*. Cambridge: Cambridge University Press, 2003; 83-99.

25. Porter TM. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press, 1995.

26. Hacking I. *The Taming of Chance*. Cambridge, UK: Cambridge University Press, 1990.

27. Krieger N. Epidemiology and the social sciences: towards a critical engagement in the 21st century. *Epidemiol Rev* 2000; 22:155-163.

28. Krieger N. Who and what is a `population'? Historical debates, current controversies, and implications for understanding `population health' and rectifying health inequities. *Milbank Quarterly* 2012; 90:634-681.

29. Hamlin C. *Public Health and Social Justice in the Age of Chadwick: Britain, 1800-1854*. Cambridge, UK: Cambridge University Press, 1998.

30. Glymour MM, Greenland S. Causal diagrams. In: Rothman K, Greenland S, Lash T (eds). *Modern Epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2008; 183-209.

31. Forita R, Spallek J, Zeeb H. Directed acyclic graphs. in: Ahrens W, Pigeot I (eds). *Handbook of Epidemiology*. 2nd ed. New York: Springer Science+Business Media, 2014; 1481-1517.

32. Pearl J. *Causality – Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press, 2009.

33. Swanson SD. Communicating causality. *Eur J Epidemiol* 2015; 30:1073-1075.

34. VanderWeele TJ, Hernán MA. Causal effects and natural laws: towards a conceptualization of causal counterfactuals for nonmanipulable exposures, with application to the effects of race and sex. In: Berzuini C, Dawid P, Bernardinelli L (eds). *Causality: Statistical Perspectives and Applications*. New York: John Wiley & Sons, Ltd., 2012; 101-113.

35. Imbens GW, Rubin DB. *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction*. Cambridge: Oxford University Press, 2015.

36. Holland PW. Statistics and causal inference. *J Am Stat Assoc* 1986; 81:945-960.

37. Curd M, Psillos S (eds). *The Routledge Companion to the Philosophy of Science*. 2nd ed. New Brunswick, NJ: Routledge, 2013.

38. Bird A, Ladyman J (eds). *Arguing about Science*. New York: Routledge, 2012.

39. Beebee H, Hitchcock C, Menzies P. *The Oxford Handbook of Causation*. Oxford: Oxford University Press, 2009.

40. Illari PM, Russo F, Williamson J (eds). *Causality in the Sciences*. New York: Oxford University Press, 2011.

41. Ziman JM. *Real Science: What it is, and What it Means*. Cambridge, UK: Cambridge University Press, 2000.

42. Lipton P. *Inference to the Best Explanation*. 2nd ed. London: Routledge, 2004.

43. Losee J. *Theories of Causality: From Antiquity to the Present*. New Brunswick, NJ: Transaction Press, 2012.

44. Reiss JR. Causation in the social sciences: evidence, inference, and purpose. *Phil Soc Sci* 2009 39:20-40.

45. Reiss J. Causation in the sciences: an inferentialist account. *Studies Hist Phil Biol Biomed Sci* 2012; 43:769-777.

46. Cartwright N. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge, UK: Cambridge University Press, 2007.

47. Cartwright N, Hardie J. *Evidence-led Policy: A Practical Guide to Doing it Better*. Oxford: Oxford University Press, 2012.

48. Davey Smith G, Hermani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics* 2014; 23(R1):R89-R98.

49. Davey Smith G. The Wright stuff: genes in the interrogation of correlation and causation. *Eur J Personality* 2012; 26:395-397.

50. Krieger N, Davey Smith G. Re: "Seeking causal explanations in social epidemiology." (letter). *Am J Epidemiol* 2000; 151:831-832.

51. Rockhill B. Theorizing about cases at the individual level while estimating effects at the population level: implications for prevention. *Epidemiol* 2005; 16:124-129.

52. Marmot M. Social causes of social inequalities in health. In: Anand S, Peter F, Sen A (eds). *Public Health, Ethics, and Equity*. New York: Oxford University Press, 2004; 37-62.

53. World Health Organization, Commission on the Social Determinants of Health. *Closing the Gap in a Generation: Health Equity Through Action on the Social Determinants of Health*. Geneva: World Health Organization, 2008. Available at: http://www.who.int/social_determinants/thecommission/finalreport/en/ ; accessed: March 23, 2016.

54. Birn AE, Pillay Y, Holtz TH. *Textbook of International Health: Global Health in a Dynamic World*. 3rd ed. New York: Oxford University Press, 2009.

55. World Health Organization. *World Health Statistics 2014*. Geneva: World Health Organization, 2014. Available at: http://www.who.int/gho/publications/world_health_statistics/2014/en/ ; accessed: March 23, 2016.

56. Global Health Watch. *Global Health Watch 4: An Alternative World Health Report*. London: Zed Books, 2014.

57. Intergovernmental Panel on Climate Change (IPCC). Fifth Assessment Report (AR5). Available at: http://ipcc.ch/report/ar5/index.shtml ; accessed: March 23, 2016.

58. McMichael AJ. Population health in the Anthropocene: gains, losses, and emerging trends. *Anthropocene Review* 2014; 1:44-56.

59. Patz JA, Frumkin H, Holloway T, Vimont DJ, Haines A. Climate change: challenges and opportunities for global health. *JAMA* 2014; 312:1565-1580.

60. UNAIDS. An Introduction to Triangulation. Geneva, Switzerland: UNAIDS, 2010. Available at: http://social3.org/a/an-introduction-to-triangulation---unaids-w25562 ; accessed: March 23, 2016.

61. Rose S. *Lifelines: Biology Beyond Determinism*. Oxford: Oxford University Press, 1998.

62. Krieger N. Embodiment: a conceptual glossary for epidemiology. *J Epidemiol Community Health* 2005; 59:350-355.

63. Hu FB. *Obesity Epidemiology*. New York: Oxford University Press, 2008.

64. Banack HR, Kaufman J. The obesity paradox: understanding the effect of obesity on mortality among individuals with cardiovascular disease. *Prev Med* 2014; 62:96-102.

65. Nestle M. Food Politics: *How the Food Industry Influences Nutrition and Health. Revised and expanded 10th anniversary edition*. Berkeley, CA: University of California Press, 2013.

66. Booth S, Coveney J. *Food Democracy: From Consumer to Food Citizen*. Springer Briefs in Public Health. Singapore: Springer, 2015.

67. Varbo A, Benn M, Davey Smith G, Timpson NJ, Tybjaerg-Hansen A, Nordestgaard BG. Remnant cholesterol, low-density lipoprotein cholesterol, and blood pressure as mediators from obesity to ischemic heart disease. *Circulation Res* Published online before print November 19, 2014, doi: 10.1161/CIRCRESAHA.116.304846

68. Würtz P, Wang Q, Kangas AJ, Richmond RC, Skarp J, Tiainen M, Tynkkynen T, Soininen P, Havulinna AS, Kaakinen M, Viikari JS, Savolainen MJ, Kähönen, M, Lehtimäki T, Männistö S, Blankenberg S, Zeller T, Laitinen J, Pouta A, Mäntyselkä P, Vanhala M, Elliott P, Pietiläinen KH, Ripatti S, Salomaa V, Raitakari OT, Järvelin M-R, Davey Smith G, Ala-Korpela M. Metabolic signatures of adiposity in young adults: mendelian randomization analysis and effects of weight change. *PLoS Med* 2014; 11(12): e1001765. doi:10.1371/journal.pmed.1001765

69. Crookshank FG. First principles: and epidemiology. *Proc Roy Soc Med* 1919-1920; *Sect Epidem*, 13:159-184.

70. Lakoff A. Vaccine politics and the management of reason. *Public Culture* 2015; 27:419-425.

71. Ganapathiraju PV, Mossink CB, Plumb J. Endgame for polio eradication? Options for overcoming social and political factors in the progress to eradicating polio. *Global Public Health* 2015; 10:463-473.

72. Lillviv DF, Kirkland A, Frick A. Power and persuasion in the vaccine debates: an analysis of political efforts and outcomes in the United States. *Milbank Q* 2014; 92:475-508.

73. Casper MJ, Carpenter LM. Sex, drugs, and politics: the HPV vaccine for cervical cancer. *Sociol Health Illness* 2008; 30:886-899.

74. PA Consulting Group. Afghanistan Stability / COIN Dynamics. Working Draft – V3. Available at: http://msnbcmedia.msn.com/i/MSNBC/Components/Photo/2009/December/091202/091203-engel-big-9a.jpg ; accessed: March 23, 2016.

75. Bummller E. We have met the enemy and he is power point. New York Times, April 26, 2010. Available at: http://www.nytimes.com/2010/04/27/world/27powerpoint.html?_r=0 ; accessed: March 23, 2016.

76. Lipton P. Précis of Inference to the Best Explanation, 2nd ed. *Philosoph Phenom Res* 2007; 74:421-423.

77. Lipton P. Causation and explanation. In: Beebee H, Hitchcock C, Menzies P. *The Oxford Handbook of Causation*. New York: Oxford University Press, 2009 ; 619-631.

78. Douven I. Abduction. *The Stanford Encyclopedia of Philosophy* (Spring 2011 edition; Zalta EN, ed). Available at: http://plato.stanford.edu/archives/spr2011/entries/abduction/ ; accessed: March 23, 2016.

79. Psillos S. The fine structure of inference to the best explanation. *Philosoph Phenom Res* 2007; 74:441-448.

80. Broadbent A. What could possibly go wrong? – a heuristic for predicting population health outcomes of interventions. *Prev Med* 2011; 53:256-259.

81. Davey Smith G, Egger M. Incommunicable knowledge? Interpreting and applying the results of clinical trials and meta-analysis. *J Clin Epidemiol* 1998; 51:289-295.

82. Koons R. Defeasible reasoning. In: Zalta EN (ed). *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). Available at: http://plato.stanford.edu/entries/reasoning-defeasible/ ; accessed: March 23, 2016.

83. Koch E, Otarala A, Kirschbaum A. A landmark for Popperian epidemiology: refutation of the randomized Aldactone evaluation study. *J Epidemiol Community Health* 2005; 59:1000-1006.

84. Hyams KC. The investigation fatigue syndrome: a case study of the limitations of inductive inferences and non-falsifiable hypotheses in medical research. *Medical Hypotheses* 2003; 60:760-766.

85. Maclure M. Multivariate refutation of etiologic hypotheses in non-experimental epidemiology. *Int J Epidemiol* 1990; 19:782-787.

86. Buck C. Popper's philosophy for epidemiologists. *Int J Epidemiol* 1975; 4:159-168.

87. Susser M. The logic of Sir Karl Popper and the practice of epidemiology. *Am J Epidemiol* 1986; 124:711-718.

88. Jacobsen M. Against Popperized epidemiology. *Int J Epidemiol* 1976; 5:9-11.

89. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd edition. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008; pp. 20-25, 30.

90. Greenland S. Induction versus Popper: substance versus semantics. *Int J Epidemiol* 1998; 4:543-548.

91. MacMahon B, Sawa JM. Physical damage to the fetus. *Milbank Quarterly* 1961; 39:14-83.

92. Denda M. Newly discovered olfactory receptors in epidermal keratinocytes are associated with proliferation, migration, and re-epithelialization of keratinocytes. *J Invest Dermatol* 2014; 134:2677-2679.

93. Busse D, Kudella P, Grüning N-M, Gisselmann G, Ständer S, Luger T, Jacobsen F, Steinsträßer L, Ralf R, Gkogkolou P, Böm M, Hatt H, Benecke H.A synthetic sandalwood odorant induces wound-healing processes in human keratinocytes via the olfactory receptor OR2AT4. *J Invest Dermatol* 2014; 134:2823-2832.

94. Oudshoorn N. *Beyond the Natural Body: An Archaeology of Sex Hormones*. London: Routledge, 1994.

95. Krieger N, Löwy I, and the "Women, Hormones, and Cancer" group (Aronowitz R, Bigby J, Dickersin K, Garner E, Gaudillière J-P, Hinestrosa C, Hubbard R, Johnson PA, Missmer SA, Norsigian J, Pearson C, Rosenberg CE, Rosenberg L, Rosenkrantz BG, Seaman B, Sonnenschein

C, Soto AM, Thorton J, Weisz G). Hormone replacement therapy, cancer, controversies & women's health: historical, epidemiological, biological, clinical and advocacy perspectives. *J Epidemiol Community Health* 2005; 59:740-748.

96. Krieger N. History, biology, and health inequities: emergent embodied phenotypes & the illustrative case of the breast cancer estrogen receptor. *Am J Public Health* 2013; 103:22-27.

97. Heldring N, Pike A, Andersson S, Matthews J, Cheng G, Hartman J, Tujague M, Ström A, Trueter E, Warner M, Gustafsson J-Å. Estrogen receptors: how do they signal and what are their targets? *Physiol Rev* 2007; 87:905-931.

98. Vrtačnik P, Ostanek B, Mencej-Bedrač S, More J. The many faces of estrogen signaling. *Biochem Med* (Zagreb) 2014; 24:329-342.

99. Terris M. (ed). *Goldberger on Pellagra*. Baton Rouge, LA: Louisiana State University Press, 1964.

100. Etheridge EW. *The Butterfly Caste: A Social History of Pellagra in the South*. Westport, CT: Greenwood Publishing Co., 1972.

101. Elmore JE, Feinstein AR. Joseph Goldberger: an unsung hero of American clinical epidemiology. *Ann Intern Med* 1994; 121:372-375.

102. Klevay LM. And so spake Goldberger in 1916: pellagra is not infectious! *J Am Coll Nutr* 1997; 16: 290-292.

103. National Institutes of Health, Office of NIH History. Dr. Joseph Goldberger and the War on Pellagra. Available at: http://history.nih.gov/exhibits/goldberger/index.html ; accessed: March 23, 2016.

104. Lansha DJ. The discovery of niacin, biotin, and panthothenic acid. *Ann Nutr Metabl* 2012; 61:246-253.

105. Davenport CB. The hereditary factor in pellagra. *Archives Intern Med* 1916; 18:4-31.

106. Goldberger J, Wheeler GA, Sullivan MX, Jones KK. *The Experimental Production of Pellagra in Human Subjects by Means of Diet*. Washington, DC: US Government Printing Office, 1920.

107. Goldberger J, Wheeler GA, Lillie RD, Rogers LM. A study of the blacktongue-preventive action of 16 foodstuffs, with special reference to the identity of blacktongue of dogs and pellagra of man. *Public Health Reports* 1928; 43:1385-1454.

108. Goldberger J, Wheeler GA, Sydenstricker E. A study of the relation of family income and other economic factors to pellagra incidence in seven cotton-mill villages of South Carolina in 1916. *Public Health Reports* 1920; 35:2673-2714.

109. Goldberger J, Wheeler GA, Sydenstricker A. A study of the relation of diet to pellagra incidence in seven textile-mill communities of South Carolina in 1916. *Public Health Reports* 1920; 35:648-713.

110. Goldberger J, Sydenstricker E. Pellagra in the Mississippi Flood Area: report of an inquiry relating to the prevalence of pellagra in the area affected by the overflow of the Mississippi and its tributaries in Tennessee, Arkansas, Mississippi, and Louisiana in the spring of 1927. *Public Health Reports* 1927; 42:2706-2725.

111. Ebrahim S. Yerushalmy and the problems of causal inference. *Int J Epidemiol* 2014; 43:1349-1351.

112. Parascandola M. Commentary: smoking, birthweight, and mortality: Jacob Yerushalmy on self-selection and the pitfalls of causal inference. *Int J Epidemiol* 2014; 43:1373-1377.

113. Yerushalmy J. The relationship of parents' cigarette smoking to outcome of pregnancy – implications to the problem of inferring causation from an observed association. *Am J Epidemiol* 1971; 93:443-456. (Reprinted *Int J Epidemiol* 2014; 43:1355-1366).

114. VanderWeele TJ. Commentary: resolution of the birthweight paradox: competing explanations and analytical insights. *Int J Epidemiol* 2014; 43:1368-1373.

115. Kramer MS, Zhang X, Platt RW. Commentary: Yerushalmy, maternal cigarette smoking and the perinatal mortality crossover. *Int J Epidemiol* 2014; 43:1378-1381.

116. Hernández-Díaz S, Schisterman EF, Hérnan MA. The birth-weight paradox uncovered? *Am J Epidemiol* 2006; 164:1115-1120.

117. VanderWeele TJ, Mumford SL, Schisterman EF. Conditioning on intermediates in perinatal epidemiology. *Epidemiol* 2012; 23:1-9.

118. Keyes KM, Davey Smith G, Susser E. Commentary: smoking in pregnancy and offspring health: early insights into family-based and 'negative-control' studies? *Int J Epidemiol* 2014; 43:1381-1388.

119. Smith GD. Assessing intrauterine influences on offspring health outcomes: can epidemiological studies yield robust outcomes? *Basic Clin Pharmacol Toxicol* 2008; 102:245-256.

120. Lowe CR. Effect of mothers' smoking habits on birth weight of their children. *Br Med J* 1959; 2:673-676.

121. Kuja-Halkola R, D'Onofrio BM, Larsson H, Lichtenstein P. Maternal smoking during pregnancy and adverse outcomes in offspring: genetic and environmental sources of covariance. *Behav Genet* 2014; 44:456-467.

122. Magnus P, Berg K, Bjerkedal T, Nance W. The heritability of smoking behavior in pregnancy, and the birth weights of offspring of smoking-discordant twins. *Scand J Soc Med* 1985; 13:29-34.

123. Tyrell J, Huikari V, Christie JT, Cavadino A, Bakker R, Brion MJ, et al. Genetic variation in the 15q25 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) interacts with maternal self-reported smoking status during pregnancy to influence birth weight. *Hum Mol Genet* 2012; 21:5344-5358.

124. Evans WN, Ringel JS. Can higher cigarette taxes improve birth outcomes? *J Public Econ* 1999; 72:135-154.

125. Hamilton BH. Estimating treatment effects in randomized clinical trials with non-compliance: the impact of maternal smoking on birthweight. *Health Econ* 2001; 10:399-410.

126. Elderton M, Pearson K. *A First Study of the Influence of Parental Alcoholism on the Physique and Ability of the Offspring*. London; Dulau and Co., 1910.

127. Pearson K, Elderton Em. *A Second Study of the Influence of Parental Alcoholism on the Physique and Ability of the Offspring*. London: Dulau and Co., 1910.

128. Stigler SM. Karl Pearson and the Cambridge economists. In: Stigler SM. *Statistics on the Table*. Cambridge, MA: Harvard University Press, 1999; 13-50.

129. Pigou AC. *Westminster Gazette*, Feb 2, 1911, pp. 1-2.

130. Davey Smith G. Pigou, Pearson, and parental intemperance: debating "collider bias" in 1911. (in preparation).

131. Berkson J. Limitations of the application of the fourfold table analysis to hospital data. *Biomet Bull* 1946; 2:47-53.

132. Greenland S, Neutra RR. An analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. *J Chron Dis* 1981; 34:433-438.

133. Greenland S. Randomization, statistics, and causal inference. *Epidemiol* 1990; 1:421-429.

134. Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. *Eur J Epidemiol* 2015; DOI: 10.1007/s10654-015-9995-7.

135. Smith GD, Ebrahim S, Frankel S. How policy informs the evidence. *BMJ* 2001; 322:184.

136. Krieger N, Alegría M, Almeida-Filho N, Barbosa da Silva J Jr, Barreto ML, Beckfield J, Berkman L, Birn A-E, Duncan BB, Franco S, Garcia DA, Gruskin S, James SA, Laurell AC, Schmidt MI, Walters KL. Who, and what, causes health inequities? Reflections on emerging debates from an exploratory Latin American/North American workshop. *J Epidemiol Community Health* 2010; 64:747-749.

137. Krieger N. Discrimination and health inequities. In: Berkman L, Kawachi I, Glymour M (eds). *Social Epidemiology*. 2nd ed. New York: Oxford University Press, 2014; 63-125.

138. Krieger N, Rowley D, Herman AA, Avery B, Phillips MT. Racism, sexism and social class: implications for studies of health, disease, and well-being. *Am J Prev Med* 1993; 9(6 suppl):82-122.

139. Gee GC, Ford CL. Structural racism and health inequities: old issues, new directions. *Du Bois Rev* 2011; 8:115-132.

140. Williams DR. Miles to go before we sleep: racial inequities in health. *J Health Soc Behav* 2012; 53:279-295.

141. Metzl JM, Roberts DE. Structural competency meets structural racism: race, politics, and the structure of medical knowledge. *Virtual Mentor: AMA J Ethics* 2014; 16:675-690.

142. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiol* 2014; 25:473-454.

143. Russo F, Wunsch G, Mouchart M. Inferring causality through counterfactuals in observational studies – some epistemological issues. *Bull Sociological Methodol* 2011; 111:43-64.

144. Greiner J, Rubin D. Causal effects of perceived immutable characteristics. *Rev Ec Stat* 2011; 93:775-785.

145. Sen M, Wasow O. Race as a "bundle of sticks": designs that estimate effects of seeming immutable characteristics. *Annual Rev Polit Sci* 2016 (in press).

146. Marks J. The nature/culture of genetic facts. *Annu Rev Anthropol* 2013; 42:247-267.

147. Wailoo K, Nelson A, Lee C. *Genetics and the Unsettled Past: The Collision of DNA, Race, and History*. New Brunswick, NJ; Rutgers, 2012.

148. Fujimora JH, Bolnick DA, Rajagopalan R, Kaufman JS, Lewontin RC, Duster T, Ossorio P, Marks J. Clines without classes: how to make sense of human variation. *Social Theory* 2014; 32:208-227.

149. Koenig B, Lee S, Richardson S (eds). *Revisiting Race in a Genomic Age*. New Brunswick, NJ: Rutgers University Press, 2008.

150. Fullwiley D. The 'contemporary synthesis': when politically inclusive genomic science relies on biological notions of race. *Isis* 2014; 105:803-814.

151. Weiss KM, Lang JC. Non-Darwinian estimation: my ancestors, my genes' ancestors. *Genome Res* 2009; 19:703-710.

152. Bliss C. Defining health justice in the postgenomic era. In: Richardson SS, Hallam S (eds). *Postgenomics: Perspectives on Biology after the Genome*. Durham, NC: Duke University Press, 2015; 174-191.

153. Frank R. Back to the future? the emergence of a geneticized conceptualization of race in sociology. *Am Acad Pol Soc Sci* 2015; 661:51-64.

154. Krieger N. Re: "On the causal interpretation of race." *Epidemiol* 2014; 25:937

155. Smith JM. On the fourteenth query of Thomas Jefferson's notes on Virginia. *Anglo-African Magazine* 1859; 1:225-228.

156. Du Bois WEB. *The Health and Physique of the Negro*. Atlanta, GA: Atlanta University Press, 1906.

157. Fix ME, Struyk RJ. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, DC: Urban Institute Press, 1993.

158. Blank RM, Dabady R, Citro CF (ed). *Measuring Racial Discrimination*. Washington, DC: National Research Council, National Academy Press, 2004.

159. Reskin B. The race discrimination system. *Annu Rev Sociol* 2012; 38:17-35.

160. Marcellisi A. Is race a cause? *Phil Sci* 2013; 80:650-659.

161. Kaufman JS. Race: ritual, regression, and reality. *Epidemiol* 2014; 25:485-487.

162. Yudell M. Race *Unmasked: Biology and Race in the Twentieth Century*. New York: Columbia University Press, 2014.

163. Ernst W, Harris B (eds). *Race, Science, and Medicine, 1700-1960*. London: Routledge, 1999.

164. Tilley H. Racial science, geopolitics, and empires: paradoxes of power. *Isis* 2014; 105:773-781.

165. Anderson W. Racial conceptions in the Global South. *Isis* 2014; 105:782-792.

166. Pascoe P. *What Comes Naturally: Miscegenation Law and the Making of Race in America*. New York: Oxford University Press, 2009.

167. Khanna N. "If you're half-black, you're just black": reflected appraisals and the persistence of the one-drop rule. *Sociol Q* 2010; 51:96-121.

168. Thorton R. Tribal membership requirements and the demography of 'old' and 'new' Native Americans. *Pop Res Policy Rev* 1997; 16:33-42.

169. Romero M, Margolis E (eds). *The Blackwell Companion to Social Inequalities*. Malden, MA: Blackwell, 2005.

170. Tilly C. *Identities, Boundaries, and Social Ties*. Boulder, CO: Paradigm Publishers, 2005.

171. Chay KY, Greenstone M. The convergence in black-white infant mortality rates during the 1960's. *Am Econ Rev.* 2000;90:326–332.

172. Almond DV, Chay KY, Greenstone M. Civil Rights, the War on Poverty, and Black-White convergence in infant mortality in the rural South and Mississippi. 31 December 2006. *MIT Economics Working Paper No. 07-04*. Available at:

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=961021 ; accessed: March 23, 2016.

173. Almond D, Chay KY. The long-run and intergenerational impact of poor infant health: evidence from cohorts born during the Civil Rights era. *Working Paper*. 2008. Available at: http://users.nber.org/~almond/chay_npc_paper.pdf ; accessed: March 23, 2016.

174. Kaplan G, Ranjit N, Burgard S. Lifting gates, lengthening lives: did civil rights policies improve the health of African-American women in the 1960s and 1970s? In: Schoeni RF, House JS, Kaplan G, Pollack H, eds. *Making Americans Healthier: Social and Economic Policy as Health Policy*. New York, NY: Russell Sage Foundation, 2008; 145–170.

175. Krieger N, Chen JT, Coull B, Waterman PD, Beckfield J. The unique impact of abolition of Jim Crow laws on reducing inequities in infant death rates and implications for choice of comparison groups in analyzing societal determinants of health. *Am J Public Health*. 2013; 103:2234–2244.

176. Krieger N, Chen JT, Coull BA, Beckfield J, Kiang MV, Waterman PD. Jim Crow and premature mortality among the US black and white population, 1960-2009: an age-period-cohort analysis. *Epidemiol* 2014; 25:494-504.

177. Beckfield J, Krieger N. Epi + demos + cracy: linking political systems and priorities to the magnitude of health inequities – evidence, gaps, and a research agenda. *Epidemiol Review* 2009; 31:152-177.

178. Thomas KK. *Deluxe Jim Crow: Civil Rights and American Health Policy, 1935-1954.* Athens, GA: University of Georgia Press, 2011.

179. de Shazo RD, Smith R, Skipworth LB. Black physicians and the struggle for civil rights: lessons from the Mississippi experience. Part 1: the forces for and against change. *Am J Med* 2014; 127:920-925.

180. Dittmer J. *The Good Doctors: The Medical Committee for Human Rights and the Struggle for Social Justice.* New York: Bloomsbury, 2010.

181. Oakes JM, Andrade KN. Methodological innovation and advances in social epidemiology. *Curr Epidemiol Rep* 2014; 1:38-44.

182. Harper S, Strumpf EL. Commentary: social epidemiology: questionable answers and answerable questions. *Epidemiol* 2012; 23:795-798.

183. Nandi A, Harper S. How consequential is social epidemiology. *Curr Epi Reports* 2014; 2:61-70.

184. Krieger N. Epidemiology and the web of causation: has anyone seen the spider? *Soc Sci Med* 1994; 39:887-903.

185. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. *Int J Epidemiol* 2001; 30:668-677.

186. Peters BG. *Strategies for comparative research in political science*. Houndmills, UK: Palgrave MacMillan 2013.

187. Rutherford GW, McFarland W, Spindler H, White K, Patel SV, Aberle-Grasse J, Sabin K, Smith N, Taché S, Calleja-Garcia JM, Stoneburner RL. Public health triangulation: approach and application to synthesizing data to understand national and local HIV epidemics. *BMC Public Health* 2010, 10:447. DOI: 10.1186/1471-2458-10-477

188. Brion MJ, Lawlor DA, Matijasevich A, Horta B, Anselmi L, Araújo CL, Menezes AM, Victora CG, Smith GD. What are causal effects of breastfeeding on IQ, obesity and blood pressure? Evidence from comparing high-income with middle-income cohorts. *Int J Epidemiol* 2011; 40:670-680.

189. Krieger N, Kosheleva A, Waterman PD, Chen JT, Beckfield J, Kiang MV. 50-year trends in US socioeconomic inequalities in health: US-born black and white Americans, 1959-2008. *Int J Epidemiol* 2014; 43:1294-131.

190. Chen JT, Beckfield J,Waterman PD, Krieger N. Can changes in the distribution of and associations between education and income bias temporal comparisons of health disparities? – an exploration with causal graphs and simulations. *Am J Epidemiol* 2013; 177:870-881.

191. Ebi KL, Rocklöv J. Climate change and health modeling: horses for courses. *Global Health Action* 2014; 7:24154. http://dx.doi.org/10.3402/gha.v7.242154

192. Klein N. *This Changes Everything: Capitalism vs The Climate*. New York: Simon & Schuster, 2014.

Table 1. Philosophical definitions of abduction, defeasible reasoning, and inference to the best explanation – and contrast to deductive reasoning and Bayesian confirmation theory.	
Abduction	In contrast to deductive reasoning, in which the inferences are necessarily true if the premises are true, the reasoning involved in induction and abduction are both "*ampliative*, meaning that the conclusion goes beyond what is (logically) contained in the premises (which is why they are non-necessary inferences) … in abduction there is an implicit or explicit appeal to explanatory considerations ... in induction, there is *only* an appeal to observed to observed frequencies or statistics" [78]. "Abduction … assigns a confirmation-theoretic role to explanation: explanatory considerations contribute to making some hypotheses more credible, and others less so. By contrast, Bayesian confirmation theory makes no references at all to the concept of explanation" [78].
Defeasible reasoning	"Reasoning is *defeasible* when the corresponding argument is rationally compelling but not deductively valid … the relationship of support between premises and conclusion is a tentative one, potentially defeated by additional information" [82]. "Philosophers David M. Armstrong and Nancy Cartwright have argued that the actual laws of nature are *oaken* rather than *iron* (to use Armstrong's terms). Oaken laws admit of exceptions: they have tacit *ceteris paribus* (other things being equal) or *ceteris absentibus* (other things being absent) conditions. As Cartwright points out, an inference based on such a law of nature is always defeasible, since we may discover additional phenomenological factors that must be added to the law in question … we know that there are many exceptional cases that we have not yet encountered and may not be able to imagine. Defeasible laws enable us to express what we really know to be the case, rather than force us to pretend that we can make an exhaustive list of all the possible exceptions … defeasible reasoning is crucial to the understanding of scientific research programs" [82].
Inference to the best explanation	"IBE is the mode of inference that proceeds as follows. D is a collection of data (facts, observations) H explains D. (H would, if true, explain D.) No other hypothesis can explain D as well as H does. _____ Therefore, H is (probably) true" [79, pp. 442-443]. "Inference to the best explanation, or `abduction' as it is sometimes called, can be seen as the extension of the idea of a self-evidencing explanation, where the phenomenon that is explained in turn provides an essential part of the reason for believing the explanation is correct … it is only by asking how well various hypotheses would explain the available evidence that we determine which hypotheses merit acceptance. In this sense, inference to the best explanation has it that explanation is prior to inference" [42, pp. 421-422]. Plausible candidates to distinguish the best explanation: "scope, precision, mechanism, unification and simplicity. Better explanations explain more types of phenomena, explain them with greater precision, provide more information about underlying mechanisms, unify apparently disparate phenomena, or simplify our overall picture of the world" [42, p. 423]. "According to the explanationist, explanatory considerations are a guide to inductive inference. We decide which of the competing hypotheses the evidence best supports by determining how well each of the competitors would explain the inference … Darwin inferred the hypothesis of natural selection because, although it was not entailed by his diverse biological evidence, natural selection would provide the best explanation of it" [42, p. 421].

Textbox 1. Crookshank FG. First principles: and epidemiology. *Proc Roy Soc Med* 1919-1920; Sect Epidem, 13:159-184. (italics in the original) [69].

pp. 178-179: "May I conclude by the brief narration of a fable? Several years ago, an ingenuous police surgeon, investigating what he was told was a case of murder, found a bullet in a heart. This he decided, and so told the coroner, was the *causa vera*, the *causa causans*, of the symptoms in this case of murder. Shortly after he went abroad to a war, and, honestly believing that war is but murder on a large scale, he investigated the appearances of many bodies; again finding bullets, he declared that bullets are the cause of war, as of murder. But, in not every fatal case was the bullet of the same kind. Moreover, the occasional absence of bullets disconcerted him until he realized that he had once found gas poisoning the *causa vera*, in a case of murder, and he therefore came to the conclusion that several wars here existed, side by side; each one *sui generis*, and boasting a different *causa vera*. He then proposed to end war by discharging other and like bullets and gases in a contrary direction, and found many who approved his plan as sensible. However, some pestilent and philosophic person told him that war was not the mere numerical exaggeration of cases of murder, brought about either by an exaltation in the virulence of bullets or gas, or by a diminution in resistance to these agencies: it was our name for a state of affairs that we conceive as brought about by the play and interplay of racial, economic, and other factors. He was told, moreover, that while undoubtedly various kinds of killing are elements of war as, conceived by the historian and statesman, wars are not to be prevented, as he hoped, by avoiding persons who, in tramcars and in cinemas, carry bullets, or who project poisonous gas in public places. He was, however, unconvinced, and returned to England more settled than ever that the causal agents of war are bullets (of various kinds, no doubt) and gases (of various toxicities, certainly), while the best hope of preventing war in future lies, not in talk about vague racial, economic, or political conditions (which can only, he thought, at most be predisposing), but in devising some means of circumventing the *causae causantes*, bullets and gas!"

Textbox 2: Inference to the Best Explanation – an exposition of Semmelweis's research programe by Peter Lipton (excerpted from Chapter 5, Contrastive inference, pp. 71-90) [42].

1) The framing of contrastive explanations: "facts vs foils"

p. 33: "What gets explained is not simply `Why this?,' but 'Why this *rather than* that?' A contrastive phenomenon consists of a fact and a foil, and the same fact may have several different foils. We may not explain why the leaves turn yellow in November *simpliciter*, but only for example why they turn yellow in November rather than in January, or why they turn yellow in November rather than blue."

p. 34: "Since the causes that explain a fact relative to one foil will not generally explain it relative to another, the contrastive question provides further restriction on explanatory causes."

pp. 36-37: "One reason that explaining a contrast is sometimes harder than explaining the fact alone is that explaining a contrast requires giving causal information that distinguishes the fact from the foil, and information that we accept as an explanation of the fact alone may not do this, since it may not include information about the foil."

2) The example of Semmelweis and explaining childbed fever

p.74: "To develop these arguments and, more generally, to show just how inferences to contrastive explanations work, it is useful to consider a simple but actual scientific example in some detail. The example I have chosen is Ignanz Semmelweis's research from 1844-8 on childbed fever … Semmelweis's central datum was that a much higher percentage of the women in the First Maternity Division of the hospital contracted the disease than in the adjacent Second Division, and Semmelweis sought to explain this difference."

p. 74: First set of hypotheses: did not mark the difference (e.g., "epidemic influence," since affected everyone, and no difference in crowding or diet while at the hospital) and so were rejected

pp. 74-75: Second set of hypotheses: did mark a difference, but did not explain it (e.g., only medical students in training treated women in the First Division and only midwives treated women in the Second Division, but both performed similar kinds of examinations, and no exams were rougher than childbirth, ruling out roughness of exam as a factor)

p. 75: Third set of hypotheses: did mark a difference, but if difference eliminated, still no effect on difference in rates (e.g., priest who delivered last rites to dying women had to pass through the First but not Second Division, suggesting that "the psychological influence of seeing the priest might explain the difference" (!), but ruled out after Semmelweis had the priest change his route and not be seen by women in either Division; also, women in First Division delivered on their backs and women in Second Division delivered on their sides, but mortality differences remained the same after Semmelweis arranged for all women to deliver on their sides)

p. 75: Final set of hypotheses: marked a difference and elimination of difference eliminated difference in mortality rates ("Kolletschka, one of Semmelweis' colleagues, received a puncture wound in his finger during an autopsy, and died from an illness with symptoms like those of childbed fever. This led Semmelweis to infer that Kolletschak's death was due to the `cadaveric matter' that the wound introduced into his own blood stream, and Semmelweis then hypothesized that the same explanation would account for deaths in the First Division, since medical students performed their examinations directly after performing autopsies, and midwives did not perform autopsies at all. Similarly, the cadaveric hypothesis would explain why women who delivered outside the hospital had a lower mortality from childbed fever, since they were not examined. Semmelweis had the medical students disinfect their hands before examination, and the mortality rate in the First Division went down to the same low level as that in the Second Division. Here at last was a difference that made a difference, and Semmelweis inferred the cadaveric hypothesis.")

Why illustration of Inference to the Best Explanation (and contrary to "hypothetico-deductive" approach):

pp. 75-76: first and second set of hypotheses rejected because although compatible with the evidence (i.e., could still be part of contributing to deaths in the First or Second Division), they could not explain the contrast between the two Divisions, nor could they explain differences observed when the "foil" changed (e.g., 'epidemic influence' could not explain why rates were higher among women in First Division as compared to women outside of the hospital)

p. 79: no expectation that the "cadaveric hypothesis" would explain ALL cases (since some women delivered by the midwives also contracted childbed fever, but the midwives had not conducted autopsies), only that it DID explain the difference between the two Divisions, a difference eliminated by disinfection after autopsy – hence "cadaveric hypothesis" incomplete, but not incorrect.

p. 80: by contrast, exposure to priest and delivery on back vs. side shown to be not only incomplete but also incorrect (since changing exposure to each made no difference to difference in mortality rates between the Divisions)

p. 81: additional "unifying" aspects of "cadaveric" hypothesis: explained both deaths of women due to childbed fever and death of the colleague, and also lower rates of mortality in women who delivered at home versus in the hospital (different foil)

p. 81: 'By tailoring his explanatory interests (and his observational and experimental procedures) to contrasts that would help discriminate between competing hypotheses, Semmelweis was able to judge which hypotheses would provide the best overall explanation of the wide variety of contrasts (and absences of contrast) he observed, and so judge which hypothesis he ought to infer. Semmelweis's inferential interests determined his explanatory interests, and the best explanation then determined his inference."

pp. 82-86: Moreover, "Inference to Best Explanation" provides a better explanation of the route to inference than the "hypothetico-deductive method" (which rejects all inductive logic) because; (a) the latter does not provide a place to begin (conjectural hypotheses are typically framed as "happy guesses"), in contrast to the clear contrastive foils used in the IBE approach; (b) Semmelweis rejected hypotheses (e.g., "epidemic influences," overcrowding) that nevertheless did not outright contradict his hypothesis and were logically compatible with it; and (c) Semmelweis accepted a hypotheses he recognized was incomplete (some women in the Second Division did die of childbed fever) but nevertheless was correct in accounting for the difference in mortality between the two Divisions.

p. 90: " … as the example shows, Inference to the Best Explanation supports a picture of research that is at once more active and realistic, where explanatory considerations guide the program of observation and experiment, as well as of conjecture. The upshot of this program is inference to the loveliest explanation but the technique is eliminative. Through the use of judiciously chosen experiments, Semmelweis determined the loveliest explanation by a process of manipulation and elimination that left only a single explanation of salient contrasts. in effect, Semmelweis converted the question of the loveliest explanation of non-contrastive facts into the question of the only explanation of various contrasts. Research programmes that make this conversion are common in science, and it is one of the merits of Inference to the Best Explanation that it elucidates this strategy."

Textbox 3. Triangulation and inference to the best explanation: the example of smoking and low birthweight.

Maternal smoking during pregnancy has been associated with offspring birthweight for many decades [120], but – as with smoking and other health outcomes – the causal nature of this association was disputed. Here we are not reviewing the (interesting) history of this debate, rather we document how findings from various approaches to strengthening causal inference can be triangulated to produce an overall evidence base that is considerably more robust than that from any individual study alone [5,119].

1) Observational studies

A large number of observational studies have demonstrated that maternal smoking during pregnancy is associated with birthweight of offspring.

2) Cross-contextual comparisons

Observational studies that have been carried out in different contexts – where the confounding of maternal smoking with socioeconomic position and related factors differ, and those that are adequately powered consistently demonstrate the same direction and approximate magnitude of association.

3) Negative control studies

Negative control studies utilising paternal smoking as a factor that may be associated with confounders to the same degree as maternal smoking, but cannot have the same magnitude of a direct intrauterine effect, demonstrate considerably larger associations of maternal than paternal smoking with offspring birthweight [119]. A second negative control – maternal smoking either before or after pregnancy but not during pregnancy – does not relate to offspring birthweight to the same extent as maternal smoking during pregnancy.

4) Within-sibship studies

Studying mothers who smoked during at least one pregnancy, and did not smoke during at least one other pregnancy, find on-average birthweight differences between the offspring born following maternal smoking and their siblings who were not exposed to antenatal smoke [121].

5) Children of twins

Offspring of female monozygotic twin pairs in which one mother smokes and the other does not smoke finds lower birthweight for the offspring of the former [122].

6) Mendelian randomization (MR)

A genetic variant which relates to heavier smoking carried by mothers is associated with lower birth weight of offspring. This association is limited to mothers who smoke, strongly suggesting that the effect of the variant is due to its influence on maternal smoking [123]. MR can be conceptualised within the instrumental variables (IV) framework. An IV is a measure that relates to the exposure of interest and is only related to the outcome of interest through this association (i.e. the IV is not associated with confounding factors and the IV has no direct influence on the outcome).

7) Non-genetic IVs

Other non-genetic IVs for maternal smoking behavior can be used. Thus a study utilized different levels of cigarette taxes across US states, which influence smoking levels, and demonstrated a birthweight-lowering effect of smoking [124].

8) Randomized controlled trials (RCTs)

In RCTs a group of mothers who are smoking before or during pregnancy are randomized to an intervention aimed at reducing smoking. Evidence that such an intervention leads to higher birthweight among offspring of the mothers randomized to the intervention has been seen [125].

Method	**Strengths in comparison to conventional observational analysis**	**Key assumptions**
Observational studies	Not applicable	No unmeasured confounders; no measurement error in assessed confounders; correctly specified model (this assumption relates to all methods to a greater or lesser degree)
C*ross-contextual comparisons*	Will reveal context-specific confounding	No unmeasured confounding which (unlike the assessed confounders) is similar in magnitude between contexts and contributes substantially to the observed associations
Negative control studies	Reveals existence of potential unmeasured confounding.	Negative control is associated with the exposure (or outcome) to the same extent as the exposure (or outcome) of interest.
Within-sibship studies	Robust to fixed maternal effects that could confound the association.	The important confounders do not change between pregnancies in a manner that is associated with change in maternal smoking behavior.
Children of twins	Between-MZ maternal twin pair analysis not subject to genetic confounding, or confounding by other factors that are shared between monozygotic twins. Comparison of between-MZ and between-DZ twin analyses allows estimation of extent of genetic confounding.	No unmeasured confounding by factors that differ between twins.
Mendelian randomization (MR)	No systematic confounding.	No pleiotropic effect of the genetic variants that influence the outcome independent of the exposure of interest.
Non-genetic IVs	No systematic confounding.	The instrumental variable does not relate to confounding factors and does not impact on the outcome except through the exposure of interest.
Randomized controlled trials (RCTs)	Randomization leads to no systematic confounding.	The intervention does not have effects except through changes in the exposure of interest.

The above is a non-exhaustive list of study designs that can contribute to triangulation of evidence. Whilst the findings of all study types can be biased, as can be seen above the source of potential bias is different across the study types, and will not associate in such a manner that possible biases would all point in the same direction (and with the same magnitude of effect) to produce the same misleading causal inference.

Textbox 4. Alternative conceptualizations of causal relationships: "race" vs racism and health.

1) Counterfactual reasoning that "race" cannot be a cause: epidemiological and social science claims

- Glass TA, Goodman SN, Hernán MA, Samet JM. Causal inference in public health. *Annu Rev Public Health* 2013; 34:61-75. [1]

  "Although the potential outcomes approach is robust in the context of a range of causal questions of high value to public health, its use raises some questions. For example, should we consider causal questions about inherent features of individuals (such as sex, race/ethnicity, or age) that cannot be reasonably translated into hypothetical interventions?" [p. 70]

- VanderWeele TJ, Whitney WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Am J Epidemiol* 2014; 25:473-454. [142]

  "Part of the challenge of interpreting race coefficients causally is that, in the formal causal inference literature, effects are often defined in terms of counterfactual or potential outcomes, which are defined as the outcomes that would result under hypothetical interventions. There are, however, no reasonable hypothetical interventions on race when race itself is the exposure." [p. 473]

  "Our discussion has focused on differences in outcomes across racial groups … A similar approach might be used with other non-manipulable exposures such as sex." [p. 480]

- VanderWeele TJ, Hernán MA. Causal effects and natural laws: towards a conceptualization of causal counterfactuals for nonmanipulable exposures, with application to the effects of race and sex. In: Berzuini C, Dawid P, Bernardinelli L (eds). *Causality: Statistical Perspectives and Applications*. New York: John Wiley & Sons, Ltd, 2012; 101-113. [34]

  "Although we believe that counterfactuals related to manipulable quantities are of primary interest for policy purposes, causation related to nonmanipulable quantities can be of scientific interest and arguably constitute a substantial portion of instances of causation in science. We consider what this approach to conceptualizing causal effects might contribute to discussion of the effects of sex and race." [p. 102]

- Russo F, Wunsch G, Mouchart M. Inferring causality through counterfactuals in observational studies – some epistemological issues. *Bull Sociological Methodol* 2011; 111:43-64. [143]

  " … attributes (such as sex or ethnicity) could indeed be considered as causes of effects, even though they cannot be manipulated physically or mentally." [p. 54]

- Greiner J, Rubin D. Causal effects of perceived immutable characteristics. *Rev Ec Stat* 2011; 93:775-785. [144]

  "The emphasis on manipulation has led some scholars (Holland, 1986a; Winship & Sobel, 1999; Freedman, 2004; Berk, 2004) to contend that it is inappropriate to conceptualize a person's actual race, sex, or national origin as a treatment in an observational study. Holland (2003) in particular distinguishes `properties' or `attributes,' such as race and sex, from 'causes,' such as a pill. The objection to studying causal effects of attributes has two aspects. Fist, attributes are not subject to change by intervention. Second, some properties (including immutable characteristics) are determined at a person's conception, and thus almost all measurable variables specific to the unit are post treatment.: `For example, because I am a White person, it would be close to ridiculous to ask what would have happened to me had I been black' (Holland, 2003)." [p. 776]

  "A shift in focus from `true' immutable characteristics to perceptions does not mean that any and all inquiries into the effect of race, sex, and so on are well defined, even those involving some aspects of randomization. Several limits are particularly important. First, if treatments are perceptions, then someone must be perceiving something." [p. 783]

2) Alternative conceptual framing of racism as a determinant of population health and health inequities

- Krieger N. Discrimination and health inequities. In: Berkman L, Kawachi I, Glymour M (eds). *Social Epidemiology*. 2nd ed. New York: Oxford, 2014; 63-125. [137]

  "To guide both the research questions posed and the methods used, ecosocial theory posits … that inequitable race relations simultaneously—and not sequentially: (*a*) benefit the groups who claim racial superiority at the expense of those whom they deem intrinsically inferior, (*b*) racialize biology to produce and justify the very

categories used to demarcate racial/ethnic groups, and (*c*) generate inequitable living and working conditions that, via embodiment, result in the biological expression of racism—and hence racial/ethnic health inequities. A corollary is that there are many pathways, not just one, by which discrimination could harm health." [p. 74]

- Gee GC, Ford Cl. Structural racism and health inequities: old issues, new directions. *Du Bois Rev* 2011; 8:115-132. [139]

  "Racial minorities bear a disproportionate burden of morbidity and mortality. These inequities might be explained by racism, given the fact that racism has restricted the lives of racial minorities and immigrants throughout history. Recent studies have documented that individuals who report experiencing racism have greater rates of illnesses. While this body of research has been invaluable in advancing knowledge on health inequities, it still locates the experiences of racism at the individual level. Yet, the health of social groups is likely most strongly affected by structural, rather than individual, phenomena. The structural forms of racism and their relationship to health inequities remain under-studied. This article reviews several ways of conceptualizing structural racism, with a focus on social segregation, immigration policy, and intergenerational effects. Studies of disparities should more seriously consider the multiple dimensions of structural racism as fundamental causes of health disparities." [p. 115]

- Williams DR. Miles to go before we sleep: racial inequities in health. *J Health Soc Behav* 2012; 53:279-295. [140]

  "Large, pervasive, and persistent racial inequalities exist in the onset, courses, and outcomes of illness. A comprehensive understanding of the patterning of racial disparities indicates that racism in both its institutional and individual forms remains an important determinant. There is an urgent need to build the science base that would identify how to trigger the conditions that would facilitate needed societal change and to identify the optimal interventions that would confront and dismantle the societal conditions that create and sustain health inequalities." [p. 279]

- Metzl JM, Roberts DE. Structural competency meets structural racism: race, politics, and the structure of medical knowledge. *Virtual Mentor: AMA J Ethics* 2014; 16:675-690. [141]

  " Physicians in the United States have long been trained to assess race and ethnicity in the context of clinical interactions. Medical students learn to identify how their patients' "demographic and cultural factors" influence their health behaviors. Interns and residents receive "cultural competency" training to help them communicate with persons of differing "ethnic" backgrounds. And clinicians are taught to observe the races of their patients and to dictate these observations into medical records—"Mr. Smith is a 45-year-old African American man"—as a matter of course.

  To be sure, attention to matters of diversity in clinical settings has been shown to affect a number of factors central to effective diagnosis and treatment. Yet an emerging educational movement challenges the basic premise that having a culturally competent or sensitive clinician reduces patients' overall experience of stigma or improves health outcomes. This movement, called "structural competency," contends that many health-related factors previously attributed to culture or ethnicity also represent the downstream consequences of decisions about larger structural contexts, including health care and food delivery systems, zoning laws, local politics, urban and rural infrastructures, structural racisms, or even the very definitions of illness and health. Locating medical approaches to racial diversity solely in the bodies, backgrounds, or attitudes of patients and doctors, therefore, leaves practitioners unprepared to address the biological, socioeconomic, and racial impacts of upstream decisions on structural factors such as expanding health and wealth disparities." [p. 674]